Youtubers Analysis

March 27, 2024

[ ]:
```
# Importing the required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

[101]:
```
#setting the style of seaborn
sns.set(style='whitegrid')
```

[102]:
```
#command to display plots inline in the notebook
%matplotlib inline
```

[103]:
```
#Ignore warnings
import warnings
warnings.filterwarnings("ignore")
```

[123]:
```
#load the dataset
youtubers1=pd.read_csv("youtubers_df.csv")
```

[124]:
```
youtubers1
```

[124]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments
0	1	tseries	Música y baile	249500000	India	86200.0	2700	78
1	2	MrBeast	Videojuegos, Humor	183500000	Estados Unidos	117400000.0	5300000	18500
2	3	CoComelon	Educación	165500000				
3	4	SETIndia	NaN	162600000				
4	5	KidsDianaShow	Animación, Juguetes	113500000				
..	…	…	…	…				
995	996	hamzymukbang	NaN	11700000				
996	997	Adaahqueen	NaN	11700000				
997	998	LittleAngelIndonesia	Música y baile	11700000				
998	999	PenMultiplex	NaN	11700000				
999	1000	OneindiaHindi	Noticias y Política	11700000				

```
2           Unknown    7000000.0    24700          0
3             India      15600.0      166          9
4           Unknown    3900000.0    12400          0
..              …           …       …         …
995  Estados Unidos     397400.0    14000        124
996           India    1100000.0    92500        164
997         Unknown     211400.0      745          0
998           India      14000.0       81          1
999           India       2200.0       31          1

                                                 Links
0    http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy…
1    http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u…
2    http://youtube.com/channel/UCbCmjCuTUZos6Inko4…
3    http://youtube.com/channel/UCpEhnqL0y41EpW2TvW…
4    http://youtube.com/channel/UCk8GzjMOrta8yxDcKf…
..                                                 …
995  http://youtube.com/channel/UCPKNKldggioffXPkSm…
996  http://youtube.com/channel/UCk3fFpqI5kDMf__mUP…
997  http://youtube.com/channel/UCdrHrQf0o0TO8YDntX…
998  http://youtube.com/channel/UC0byBrdrtQ20BU9PxH…
999  http://youtube.com/channel/UC0jgc1p2hJ4GZi6pQQ…

[1000 rows x 9 columns]
```

[125]:
```
#correcting the column header by renaming Suscribers to Subscribers
youtubers = youtubers1.rename(columns={'Suscribers': 'Subscribers'})
```

[126]:
```
youtubers
```

[126]:
```
     Rank              Username           Categories  Subscribers  \
0       1               tseries       Música y baile    249500000
1       2               MrBeast   Videojuegos, Humor    183500000
2       3             CoComelon            Educación    165500000
3       4               SETIndia                  NaN    162600000
4       5         KidsDianaShow  Animación, Juguetes    113500000
..    …                     …                  …          …
995   996          hamzymukbang                  NaN     11700000
996   997             Adaahqueen                 NaN     11700000
997   998  LittleAngelIndonesia       Música y baile     11700000
998   999          PenMultiplex                  NaN     11700000
999  1000         OneindiaHindi  Noticias y Política     11700000

            Country       Visits    Likes  Comments  \
0             India      86200.0     2700        78
1    Estados Unidos  117400000.0  5300000     18500
2           Unknown    7000000.0    24700         0
```

```
3             India      15600.0      166          9
4           Unknown    3900000.0    12400          0
..              …           …       …         …
995  Estados Unidos     397400.0    14000        124
996           India    1100000.0    92500        164
997         Unknown     211400.0      745          0
998           India      14000.0       81          1
999           India       2200.0       31          1

                                                 Links
0    http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy…
1    http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u…
2    http://youtube.com/channel/UCbCmjCuTUZos6Inko4…
3    http://youtube.com/channel/UCpEhnqL0y41EpW2TvW…
4    http://youtube.com/channel/UCk8GzjMOrta8yxDcKf…
..                                                 …
995  http://youtube.com/channel/UCPKNKldggioffXPkSm…
996  http://youtube.com/channel/UCk3fFpqI5kDMf__mUP…
997  http://youtube.com/channel/UCdrHrQf0o0TO8YDntX…
998  http://youtube.com/channel/UCObyBrdrtQ20BU9PxH…
999  http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ…

[1000 rows x 9 columns]
```

[127]:
```
#to display the first 5 rows
youtubers.head()
```

[127]:
```
   Rank       Username           Categories  Subscribers         Country  \
0     1        tseries       Música y baile    249500000           India
1     2        MrBeast   Videojuegos, Humor    183500000  Estados Unidos
2     3      CoComelon            Educación    165500000         Unknown
3     4       SETIndia                  NaN    162600000           India
4     5  KidsDianaShow  Animación, Juguetes    113500000         Unknown

        Visits    Likes  Comments  \
0      86200.0     2700        78
1  117400000.0  5300000     18500
2    7000000.0    24700         0
3      15600.0      166         9
4    3900000.0    12400         0

                                             Links
0  http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy…
1  http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u…
2  http://youtube.com/channel/UCbCmjCuTUZos6Inko4…
3  http://youtube.com/channel/UCpEhnqL0y41EpW2TvW…
4  http://youtube.com/channel/UCk8GzjMOrta8yxDcKf…
```

[128]:
```
#to display the last 5 rows
youtubers.tail()
```

[128]:

	Rank	Username	Categories	Subscribers	\
995	996	hamzymukbang	NaN	11700000	
996	997	Adaahqueen	NaN	11700000	
997	998	LittleAngelIndonesia	Música y baile	11700000	
998	999	PenMultiplex	NaN	11700000	
999	1000	OneindiaHindi	Noticias y Política	11700000	

	Country	Visits	Likes	Comments	\
995	Estados Unidos	397400.0	14000	124	
996	India	1100000.0	92500	164	
997	Unknown	211400.0	745	0	
998	India	14000.0	81	1	
999	India	2200.0	31	1	

	Links
995	http://youtube.com/channel/UCPKNKldggioffXPkSm…
996	http://youtube.com/channel/UCk3fFpqI5kDMf__mUP…
997	http://youtube.com/channel/UCdrHrQf0o0TO8YDntX…
998	http://youtube.com/channel/UCObyBrdrtQ20BU9PxH…
999	http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ…

[129]:
```
#to check the number of null values present
youtubers.isnull()
```

[129]:

	Rank	Username	Categories	Subscribers	Country	Visits	Likes	\
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	True	False	False	False	False	
4	False	False	False	False	False	False	False	
..	…	…	…	…	…	…	…	
995	False	False	True	False	False	False	False	
996	False	False	True	False	False	False	False	
997	False	False	False	False	False	False	False	
998	False	False	True	False	False	False	False	
999	False	False	False	False	False	False	False	

	Comments	Links
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
..	…	…

```
995     False  False
996     False  False
997     False  False
998     False  False
999     False  False

[1000 rows x 9 columns]
```

[130]:
```
#To check the number of null values in each column
youtubers.isnull().sum()
```

[130]:
```
Rank            0
Username        0
Categories    306
Subscribers     0
Country         0
Visits          0
Likes           0
Comments        0
Links           0
dtype: int64
```

[131]:
```
#statistical summary of the dataset
youtubers.describe()
```

[131]:

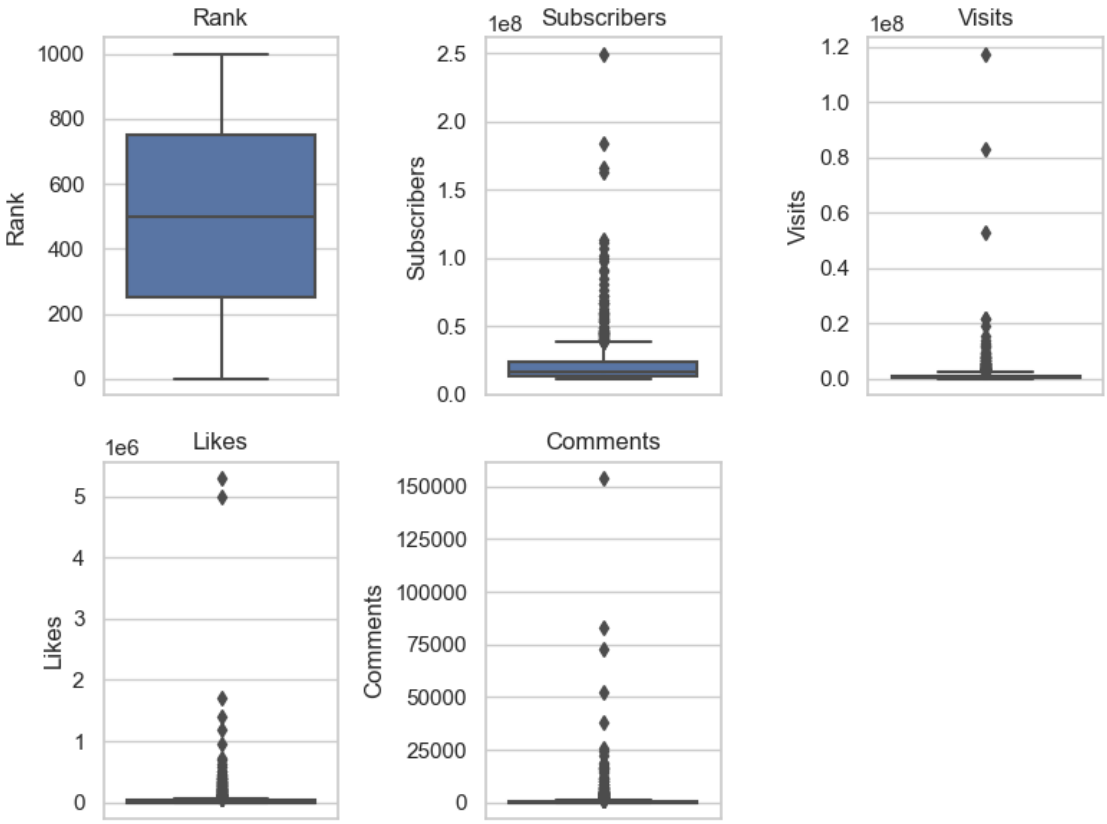
	Rank	Subscribers	Visits	Likes	Comments
count	1000.000000	1.000000e+03	1.000000e+03	1.000000e+03	1000.000000
mean	500.500000	2.189440e+07	1.209446e+06	5.363259e+04	1288.768000
std	288.819436	1.682775e+07	5.229942e+06	2.580457e+05	6778.188308
min	1.000000	1.170000e+07	0.000000e+00	0.000000e+00	0.000000
25%	250.750000	1.380000e+07	3.197500e+04	4.717500e+02	2.000000
50%	500.500000	1.675000e+07	1.744500e+05	3.500000e+03	67.000000
75%	750.250000	2.370000e+07	8.654750e+05	2.865000e+04	472.000000
max	1000.000000	2.495000e+08	1.174000e+08	5.300000e+06	154000.000000

[132]:
```
#basic strudture of the dataset
youtubers.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Rank         1000 non-null   int64
 1   Username     1000 non-null   object
 2   Categories   694 non-null    object
 3   Subscribers  1000 non-null   int64
 4   Country      1000 non-null   object
```

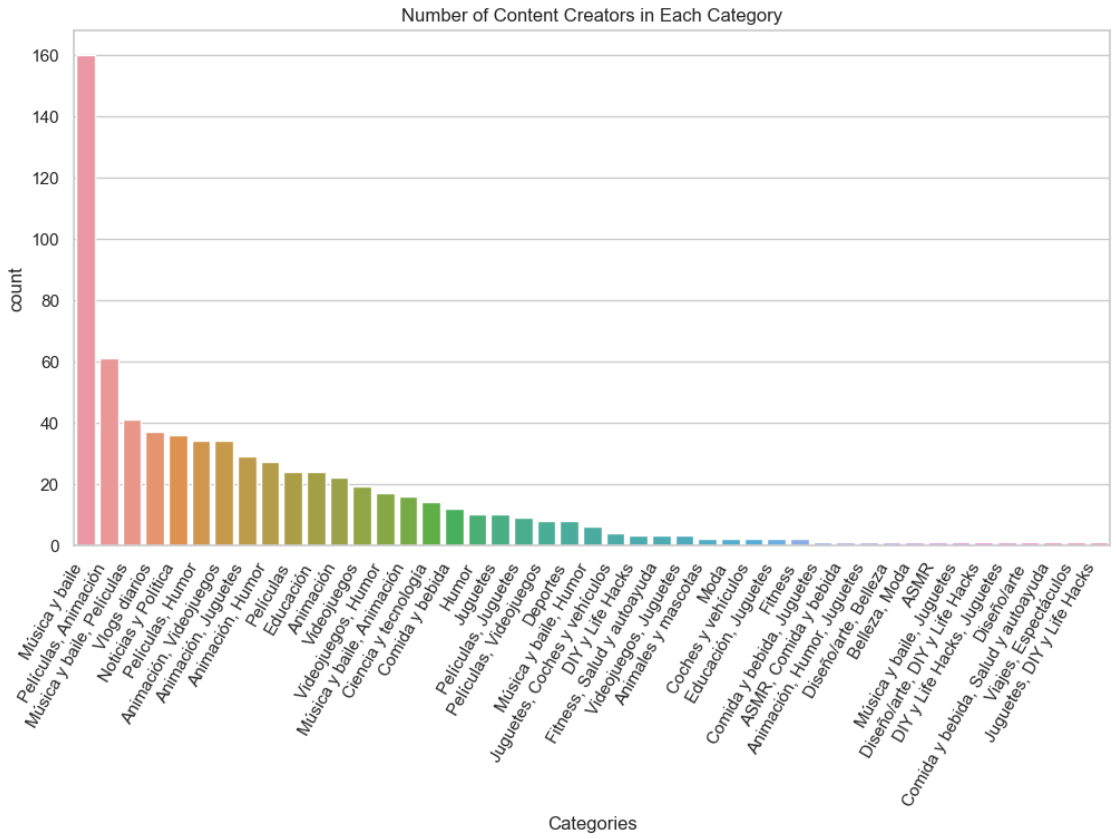
```
 5   Visits        1000 non-null   float64
 6   Likes         1000 non-null   int64
 7   Comments      1000 non-null   int64
 8   Links         1000 non-null   object
dtypes: float64(1), int64(4), object(4)
memory usage: 70.4+ KB
```

[133]:
```
#check for outliers by inspecting the dataset using box plots
numeric_columns = ['Rank','Subscribers','Visits','Likes','Comments']
plt.figure(figsize=(8, 6))
for i, column in enumerate (numeric_columns):
    plt.subplot(2,3, i+1)
    sns.boxplot(data=youtubers, y=column)
    plt.title(column)
plt.tight_layout()
plt.show()
```

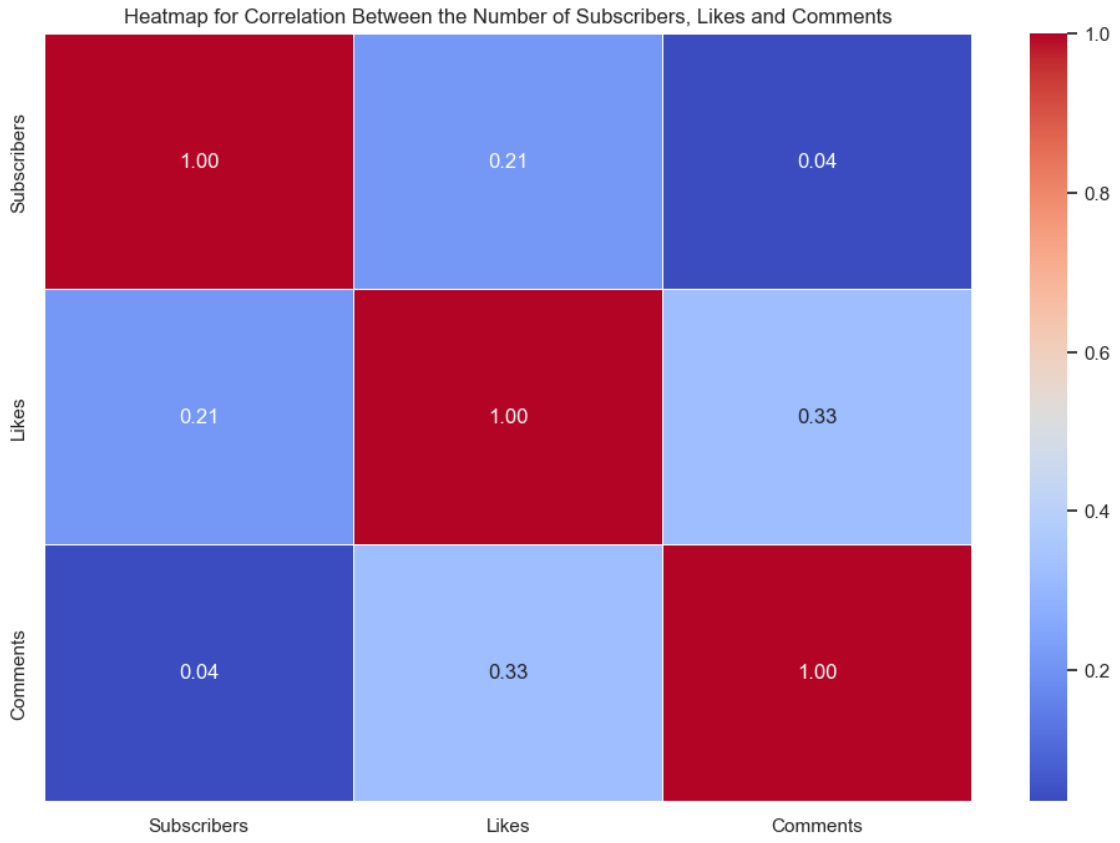


[134]:
```
#Identify the most popular categories
plt.figure(figsize=(12,6))
sns.countplot(x='Categories', data=youtubers, order=youtubers['Categories'].
 ↪value_counts().index)
```

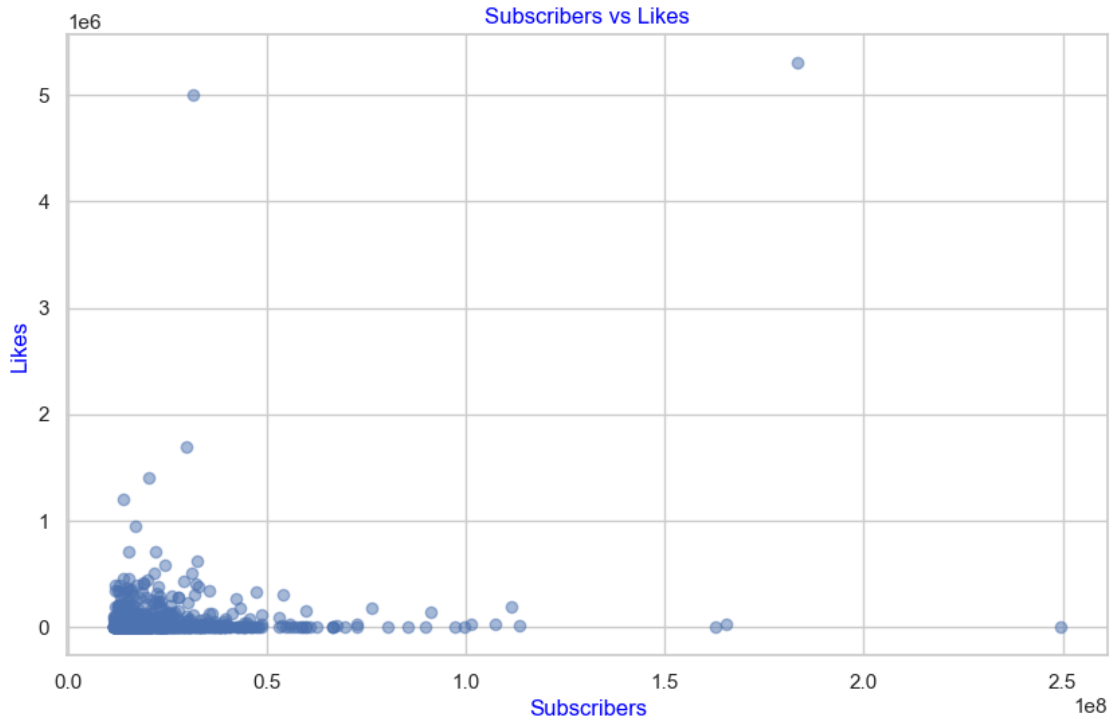
```
plt.title("Number of Content Creators in Each Category")
plt.xticks(rotation=60, ha='right')
plt.show()
```



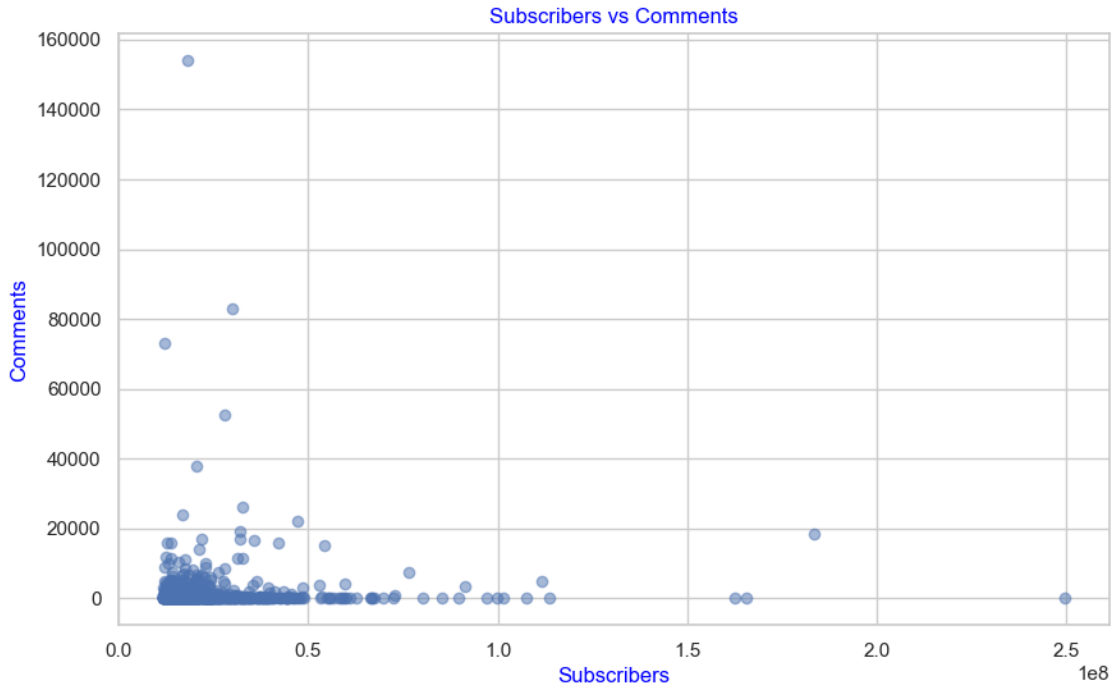
[135]:
```
#Analyzing correlation between the number of subscribers,likes, comments
correlation_yt = youtubers[['Subscribers', 'Likes','Comments']].corr()
plt.figure(figsize=(12,8))
sns.heatmap(correlation_yt, annot=True, cmap='coolwarm', fmt='.2f',linewidths=.
↪5)
plt.title("Heatmap for Correlation Between the Number of Subscribers, Likes and␣
↪Comments")
plt.show()
```



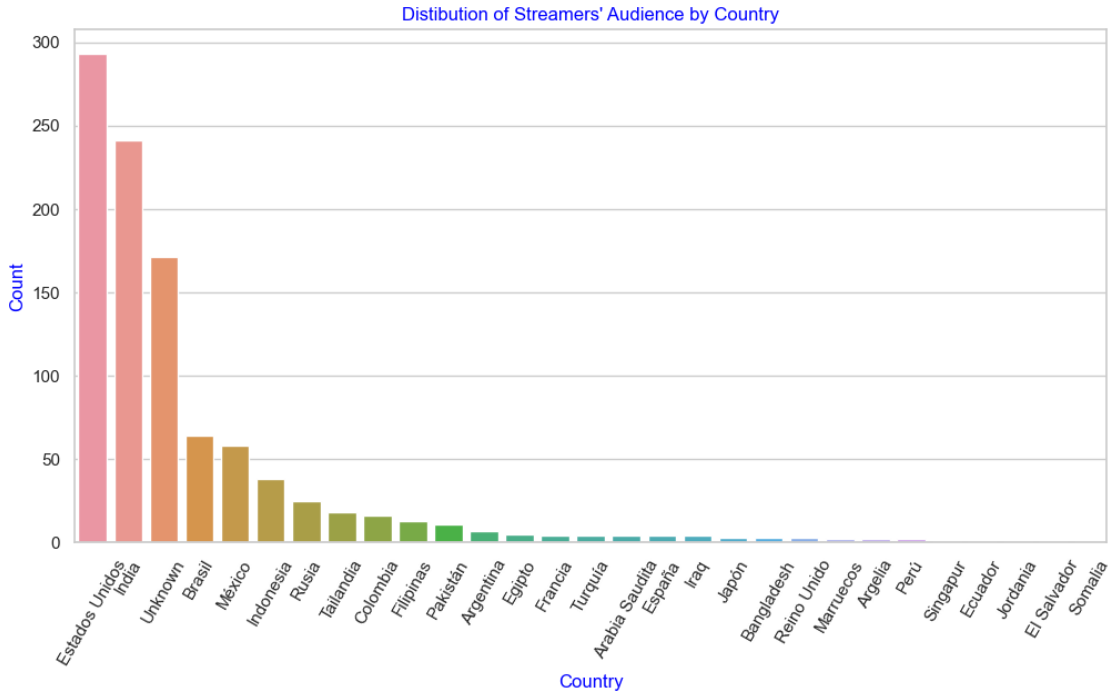
[136]:
```
#Scatter plot to display subscribers vs. Likes
plt.figure(figsize=(10, 6))
plt.scatter(youtubers['Subscribers'], youtubers['Likes'], alpha=0.5)
plt.title('Subscribers vs Likes',color='Blue')
plt.xlabel('Subscribers',color='Blue')
plt.ylabel('Likes',color='Blue')
plt.grid(True)
plt.show()
```



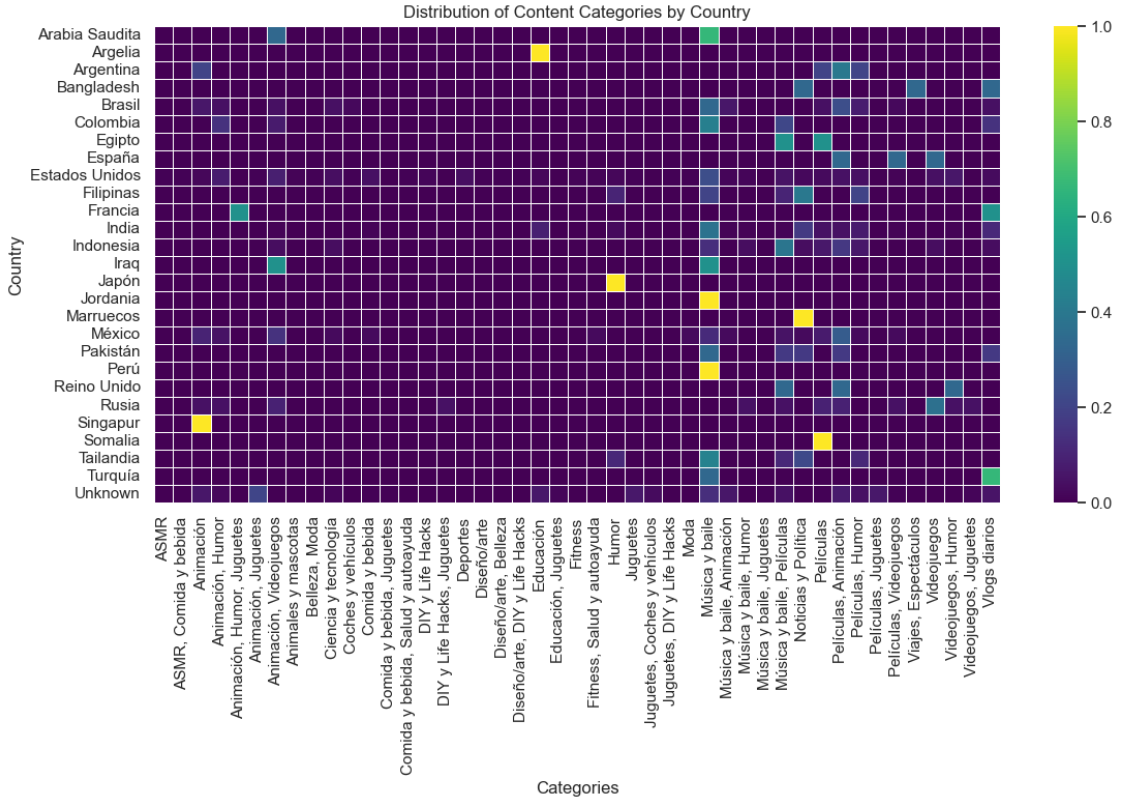
[137]:
```
#Scatter plot to display subscribers vs. comments
plt.figure(figsize=(10, 6))
plt.scatter(youtubers['Subscribers'], youtubers['Comments'], alpha=0.5)
plt.title('Subscribers vs Comments',color='Blue')
plt.xlabel('Subscribers', color='Blue')
plt.ylabel('Comments', color='Blue')
plt.grid(True)
plt.show()
```



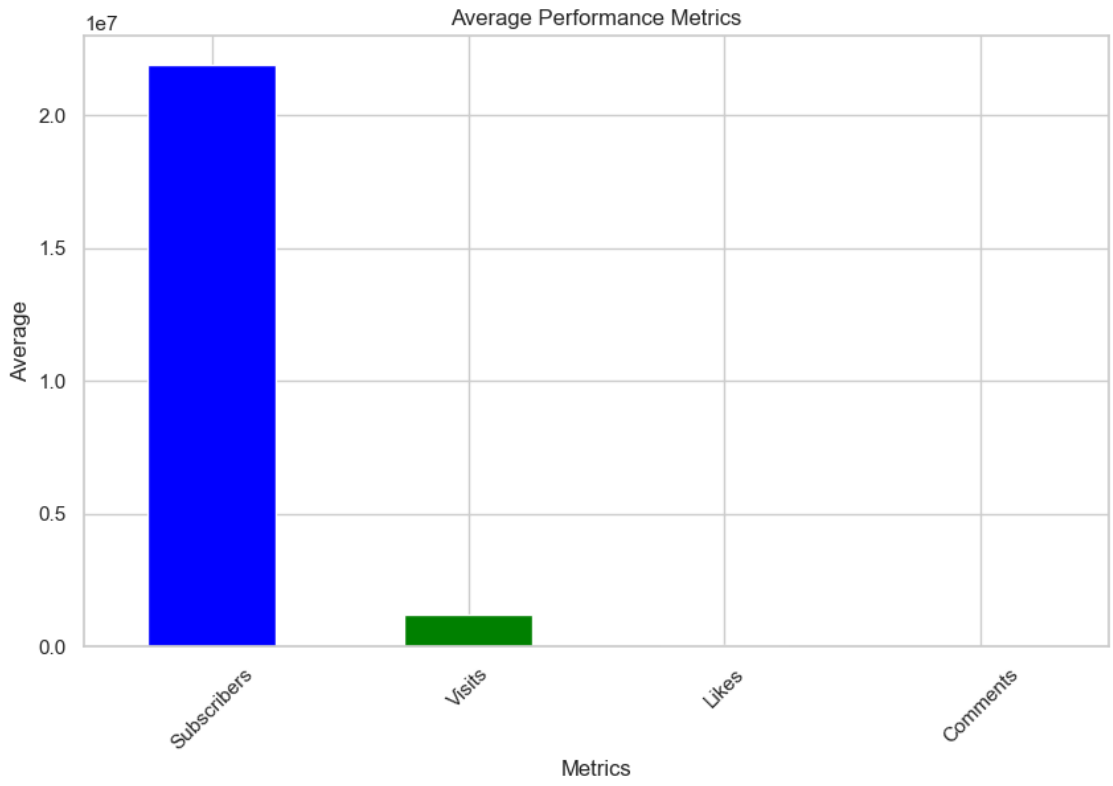
[138]:
```
#Analyze the distibution of streamers' audience by country
country_count = youtubers ['Country'].value_counts()
plt.figure(figsize=(12,6))
sns.barplot (x=country_count.index, y=country_count.values)
plt.title("Distibution of Streamers' Audience by Country",  color='Blue')
plt.xlabel('Country',  color='Blue')
plt.ylabel('Count',  color='Blue')
plt.xticks(rotation=60)
plt.show()
```



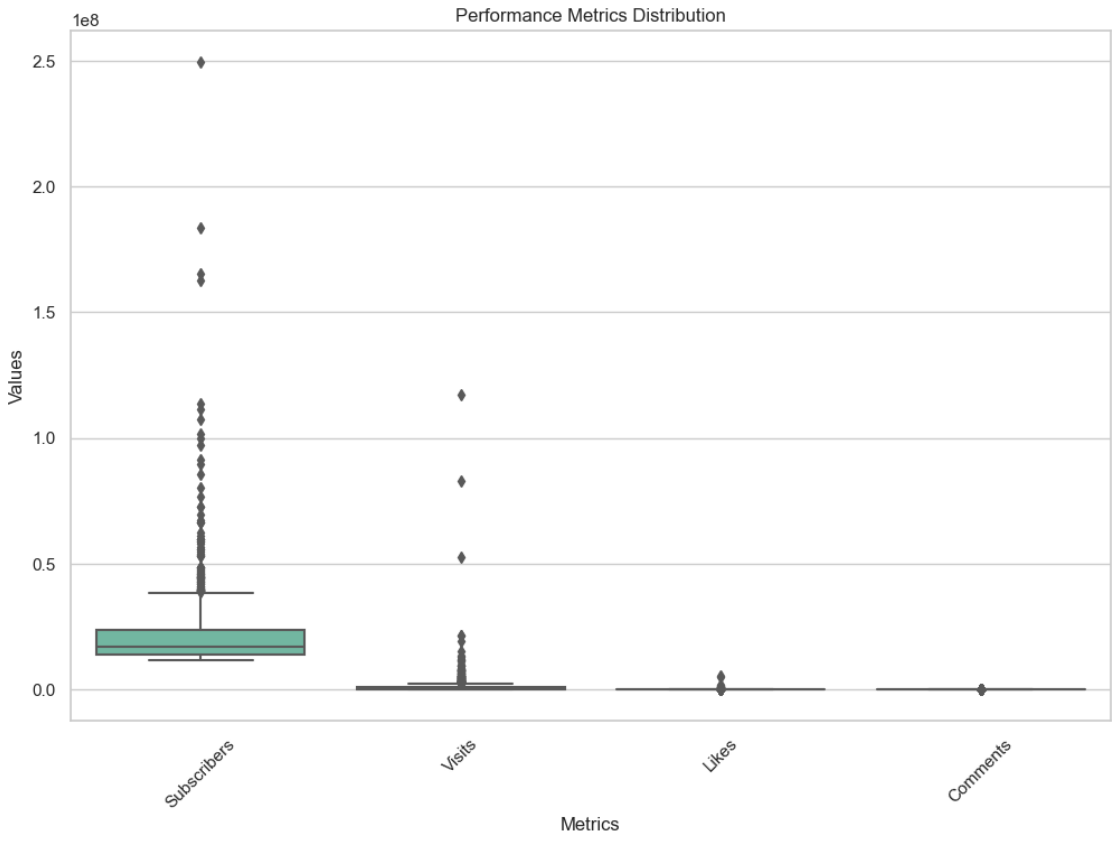
[139]:
```
#Analyze regional preferences for specific content category

# Group the data by 'Country' and 'Categories' and count the occurrences
country_categories_counts = youtubers.groupby(['Country', 'Categories']).size().
 ↪unstack(fill_value=0)

# Normalize the counts to get percentages
country_categories_perc = country_categories_counts.
 ↪div(country_categories_counts.sum(axis=1), axis=0)

# Plotting
plt.figure(figsize=(12, 8))
sns.heatmap(country_categories_perc, cmap='viridis', linewidths=.5)
plt.title('Distribution of Content Categories by Country')
plt.xlabel('Categories')
plt.ylabel('Country')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



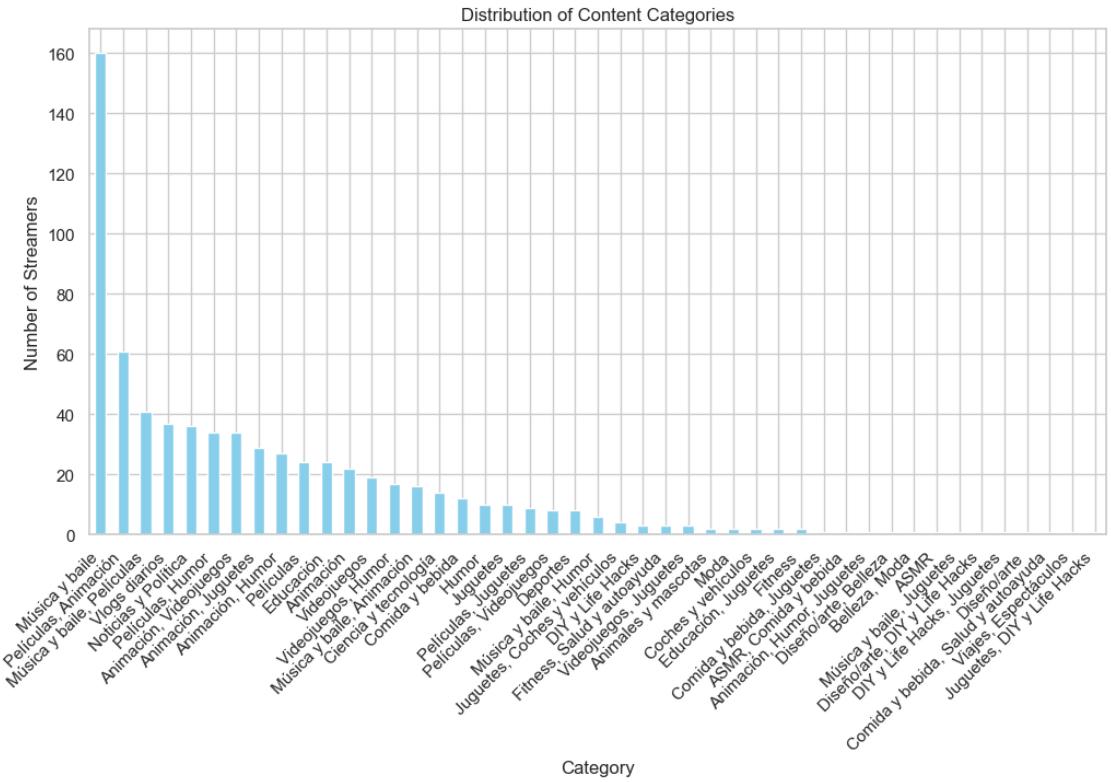
[143]:
```
#Analyzing performance metrics
# Calculate average metrics
average_metrics = youtubers[['Subscribers', 'Visits', 'Likes', 'Comments']].
 ↪mean()
# Visualize average metrics
plt.figure(figsize=(10, 6))
average_metrics.plot(kind='bar', color=['blue', 'green', 'orange', 'red'])
plt.title('Average Performance Metrics')
plt.xlabel('Metrics')
plt.ylabel('Average')
plt.xticks(rotation=45)
plt.show()
```



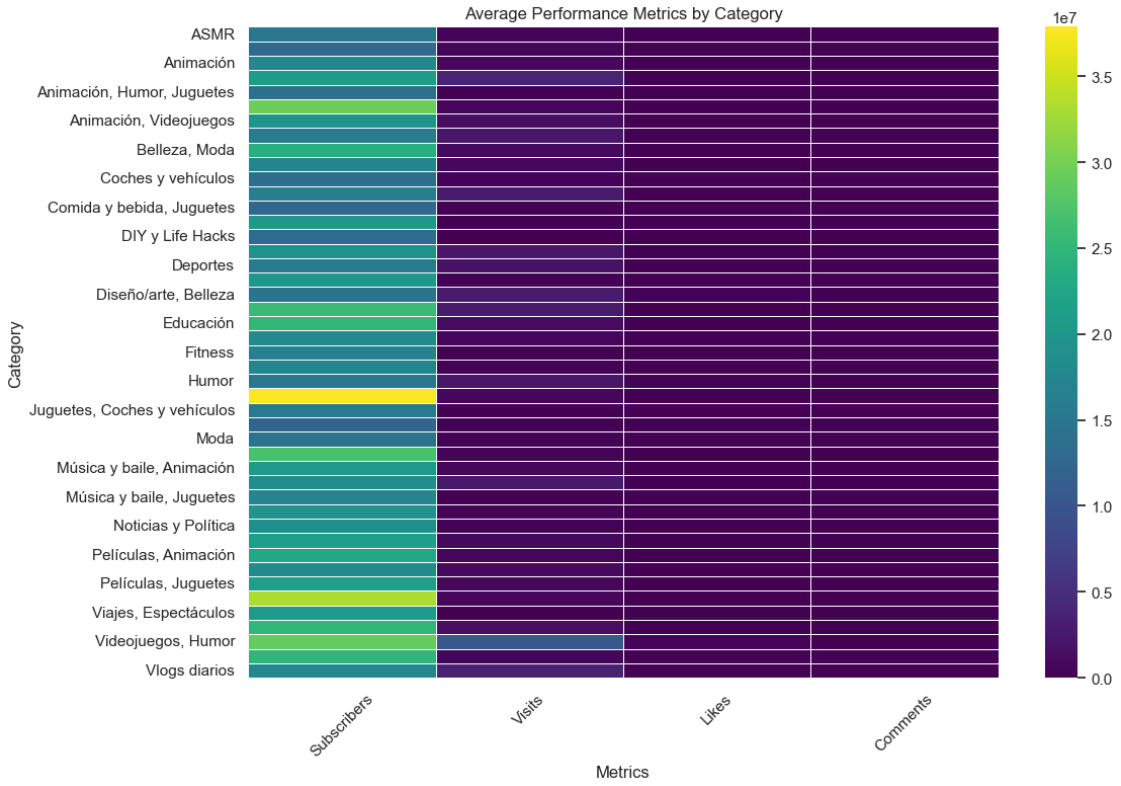
[144]:
```
# Check for patterns or anomalies
# You can use box plots to visualize the distribution of each metric
plt.figure(figsize=(12, 8))
sns.boxplot(data=youtubers[['Subscribers', 'Visits', 'Likes', 'Comments']],␣
 ↪palette='Set2')
plt.title('Performance Metrics Distribution')
plt.xlabel('Metrics')
plt.ylabel('Values')
plt.xticks(rotation=45)
plt.show()
```



[146]:
```
#Analysis of Content Categories
# Explore the distribution of content categories
category_distribution = youtubers['Categories'].value_counts()

# Plot the distribution of content categories
plt.figure(figsize=(12, 6))
category_distribution.plot(kind='bar', color='skyblue')
plt.title('Distribution of Content Categories')
plt.xlabel('Category')
plt.ylabel('Number of Streamers')
plt.xticks(rotation=45, ha='right')
plt.show()
```



[147]:
```
# Identify categories with exceptional performance metrics
# You can calculate average metrics for each category
category_performance = youtubers.groupby('Categories')[['Subscribers',␣
 ↪'Visits', 'Likes', 'Comments']].mean()

# Plot performance metrics for each category
plt.figure(figsize=(12, 8))
sns.heatmap(category_performance, cmap='viridis', linewidths=.5)
plt.title('Average Performance Metrics by Category')
plt.xlabel('Metrics')
plt.ylabel('Category')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[153]:
```
#Benchmarking analysis
# Calculate average performance metrics
average_subscribers = youtubers['Subscribers'].mean()
average_visits = youtubers['Visits'].mean()
average_likes = youtubers['Likes'].mean()
average_comments = youtubers['Comments'].mean()

# Filter streamers with above-average performance
above_average_streamers = youtubers[
    (youtubers['Subscribers'] > average_subscribers) &
    (youtubers['Visits'] > average_visits) &
    (youtubers['Likes'] > average_likes) &
    (youtubers['Comments'] > average_comments)
]

# Determine the top-performing content creators
top_performers = above_average_streamers.sort_values(by=['Subscribers',␣
 ↪'Visits', 'Likes', 'Comments'], ascending=False)

# Display the top-performing content creators
print("Top Performing Content Creators:")
```

```
print(top_performers[['Username', 'Subscribers', 'Visits', 'Likes',␣
 ↪'Comments']])
```

```
Top Performing Content Creators:
                  Username  Subscribers       Visits    Likes  Comments
1                  MrBeast    183500000  117400000.0  5300000     18500
5                PewDiePie    111500000    2400000.0   197300      4900
26             dudeperfect     59700000    5300000.0   156500      4200
34             TaylorSwift     54100000    4300000.0   300400     15000
39             JuegaGerman     48600000    2000000.0   117100      3000
43                A4a4a4a4     47300000    9700000.0   330400     22000
58               Mikecrack     43400000    2200000.0   183400      1800
62          KimberlyLoaiza     42100000    5300000.0   271300     16000
64         luisitocomunica     41100000    2500000.0   128900      1800
70             JessNoLimit     39600000    1300000.0    73500      1600
96          TotalGaming093     36300000    1500000.0   129400      4900
98    TechnoGamerzOfficial     35600000    6200000.0   341800     16500
100             markiplier     35500000    2100000.0   126500      3800
122                AboFlah     32700000    3300000.0   382000     11400
123         MRINDIANHACKER     32600000    6500000.0   617400     26000
131           fedevigevani     32000000    7700000.0   412200     17000
132                  dream     31900000    3300000.0   309200     19000
136               MrBeast2     31300000   83100000.0  5000000     11600
145          jacksepticeye     30400000    1600000.0    83400      2300
153              DaFuqBoom     29800000   52700000.0  1700000     82800
176              CrazyXYZ      27800000    4200000.0   284100      8600
177                DanTDM      27800000    3500000.0   285000     52500
179            brentrivera     27600000    6400000.0   154100      5000
180               NichLmao     27500000    1500000.0    85800      1600
195             nickiminaj     26100000    1600000.0    98300      7600
206              AlejoIgoa     25700000    5700000.0   208400      1700
207                  ZHCYT     25700000    2600000.0   127300      2200
238             alanbecker     24300000    7600000.0   582600      5900
234                    rug     24300000    3200000.0    85300      5100
241     juandediospantojaa     24000000    3000000.0   133200      3600
266           DrossRotzank     23100000    1700000.0   105900      3900
272          AmiRodrigueZZ     22900000    4300000.0   294400      1300
278            StokesTwins     22700000   11700000.0   235000     10000
282   souravjoshivlogs7028     22700000    5600000.0   382300      8900
281                SSundee     22700000    1700000.0    59800      1800
288  VillageCookingChannel     22500000   21500000.0   321500      5900
300            alfredolarin     21900000   12900000.0   707600      2100
302              royaltyfam     21900000    4700000.0    67000      6600
```

[ ]: